Journal of Theoretical Biology 393 (2016) 203–217

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Applications and implications of the exponentially modified gamma distribution as a model for time variabilities related to cell proliferation and gene expression

A. Golubev

Department of Biochemistry, Saint-Petersburg State University, Saint Petersburg, Russia

HIGHLIGHTS

- A closed form formula for convolved gamma and exponent distributions is suggested.
- It fits cell interdivision time distributions better than other functions do.
- It fits as well the distributions of times between gene expression bursts.
- Implications of fitting the two distributions with one function are discussed.

ARTICLE INFO

Article history:
Received 6 May 2015
Received in revised form
7 December 2015
Accepted 16 December 2015
Available online 15 January 2016

Keywords:
Cell cycle
Transcriptional cycle
Single cell
Time distribution
Transition probability

ABSTRACT

A panel of published distributions of cell interdivision times (IDT) comprising 77 datasets related to 16 cell types, some studied under different conditions, was used to evaluate their conformance to the exponentially modified gamma distribution (EMGD) in comparison with distributions suggested for IDT data earlier. Lognormal, gamma, inverse Gaussian, and shifted Weibull and gamma distributions were found to be generally inferior to EMGD. Exponentially modified Gaussian (EMG) performed equally well. Although EMGD or EMG may be worse than some other distributions in specific cases, the reason that IDT distributions must be generated by a common mechanism of the cell cycle makes it unlikely that they differ essentially in different cell types. Therefore, exponentially modified peak functions, such as EMGD or EMG, are most appropriate if the use of a single distribution for IDT data is reasonable. EMGD is also shown to be the best descriptive tool for published data on the distribution of times between the bursts of mRNA synthesis at defined genes in single cells. EMG is inadequate to such data because its Gaussian component markedly extends to the negative time domain. The applicability of EMGD to comparable features of cells and genes behaviors are discussed to support the validity of the transition probability model and to relate the exponential component of EMGD to the times of cell dwelling in the restriction point of the cell cycle.

1. Background

Times between consecutive cell divisions (interdivision times, IDTs) are highly variable, even in clonal cells, which may seem quite similar in all other respects. Examples in Fig. 1 show how broad IDT distributions can be. Their another noteworthy feature is positive skewness. Analytical approximations (probability density functions, PDFs) of IDT distributions are increasingly recognized as essential for models of cell and tissue kinetics (León et al., 2004; Nakaoka and Inaba, 2014), such as related to immune responses, which involve the expansion of lymphocyte clones (Duffy et al., 2012; Zilman et al., 2010), or to tumor responsiveness to therapeutic interventions as it depends on tumor cell proliferation state and history (Gabriel et al., 2012; Maler and Lutscher, 2010), and may also prompt insights into the intracellular mechanisms of the cell cycle that are responsible for the observed patterns of IDT variability. These two aspects of IDT variability suggest different criteria for choosing which model is more adequate: one showing a better formal fit to data, or one derived from a reasonable idea about cell cycle mechanism and showing an acceptable fit, even if being inferior, in this regard, to alternatives less supported theoretically.

PDFs used to approximate or model IDT variability are presented in Table 1. Their fits to empirical IDT distributions, such as exemplified in Fig. 1, are very similar and seem good enough to be

E-mail address: lxglbv@rambler.ru

http://dx.doi.org/10.1016/j.jtbi.2015.12.027
0022-5193/

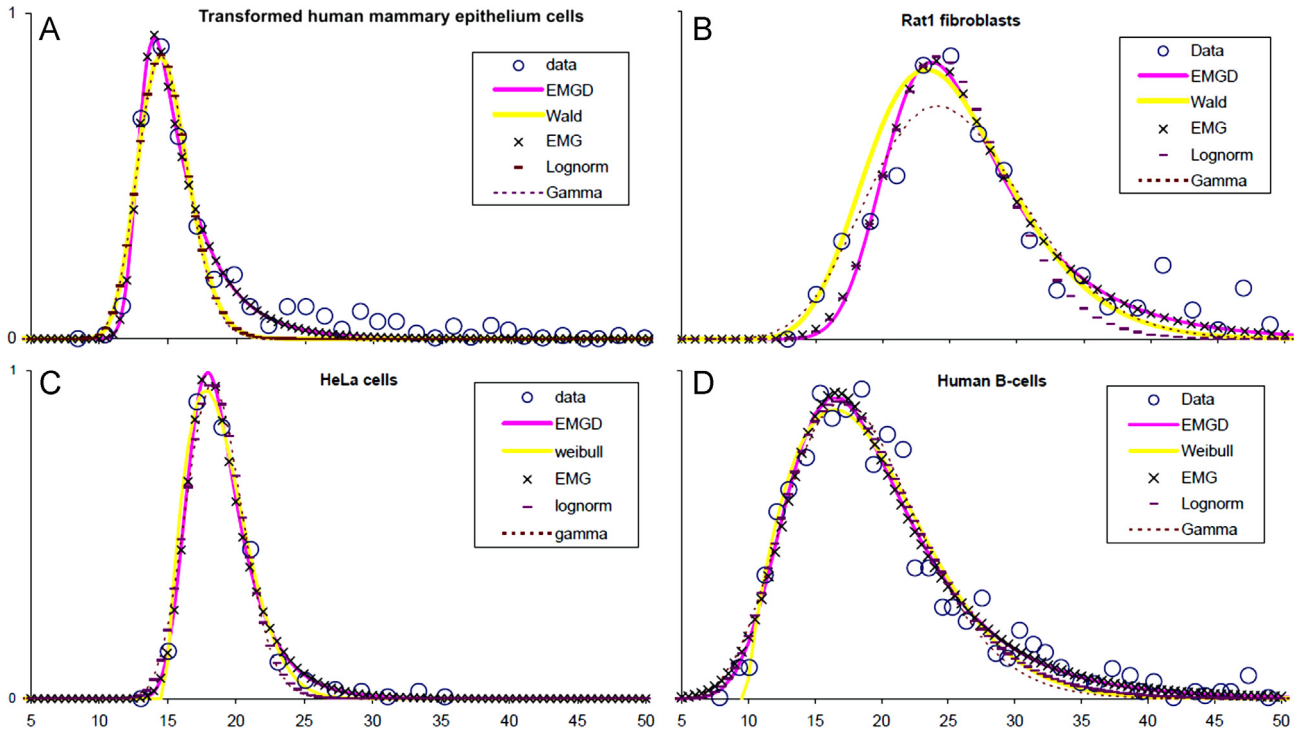

Fig. 1. Exemplary cell interdivision time distributions. Abscissas show time (hours). The centers of open circles correspond to the centers of the tops of the bars of source histograms. Because the histograms have different vertical scales, which present the number or percent of cells that fall within a time interval represented by the width of a bar, the data are aligned vertically by ascribing unity to the tops of the ordinates of source figures before being fitted with functions indicated in inserts. The plots of the functions are presented as discrete symbols spaced in 0.5 h (A, C) or 1 h (B, D) for exponentially modified Gaussian (EMG) and lognormal distributions or as smooth lines for exponentially modified gamma distribution (EMGD), inverse Gaussian (Wald) and shifted gamma (dotted line) and Weibull distributions. Data sources: A – Fig. 3c in Tyson et al. (2012); B – Fig. 1b in Holzel et al. (2001); C – Fig. 4b in Hurwitz and Tolmach (1969); D – Fig. S3 (generation 3) in Duffy et al. (2012).

Table 1
Distributions used to model or/and approximate cell interdivision time variability.

Distribution	Probability density: $\lim_{\Delta t\to 0}\left[\frac{N(t)-N(t-\Delta t)}{\Delta t}\right]=n(t)$	Mean	Variance	References* and notes
Lognormal	$\frac{1}{t\cdot\sqrt{2\pi}\cdot\sigma}\cdot\exp\left[\frac{-(\ln t-\mu)^2}{2\sigma^2}\right]$	$e^{\mu+\delta^2/2}$	$(e^{\sigma^2}-1)\cdot e^{2\mu+\sigma^2}$	1; 2; 3; 12
Gamma	$\frac{t^{c-1}\exp(-\frac{t}{b})}{b^c\Gamma(c)}$	$b\cdot c$	$b^2\cdot c$	1; 4; 5; 7; shifted form: 6;8
Wald	$\sqrt{\frac{\lambda}{2\pi\cdot t^3}}\cdot\exp\left[-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right]$	μ	μ^3/λ	3; 13; simple and shifted forms and their convolution: 9
Weibull	$\frac{b}{c}\cdot(t/c)^{b-1}\exp[-(t/c)^b]$	$c\cdot\Gamma(1+1/b)$	$c\cdot\{\Gamma(1+2/b)-\ -[\Gamma(1+1/b)]^2\}$	1
EMG	$\frac{1}{2\tau}\cdot\exp\left(\frac{2\tau(\mu-t)+\sigma^2}{2\tau^2}\right)\cdot\mathrm{erfc}\left(\frac{\tau(\mu-t)+\sigma^2}{\sqrt{2}\sigma\cdot\tau}\right)$	$\mu+\tau$	$\sigma^2+\tau^2$	Exponentially modified Gaussian: 3; 9; 10; 11
EMGD	$e^{-\frac{t}{\tau}}\cdot b^{-c}\cdot\left(\frac{\tau-b}{\tau\cdot b}\right)^{-c}\cdot[\tau\cdot\Gamma(c)]^{-1}\cdot\left[\Gamma(c)-\Gamma\left(c,\frac{\tau-b}{\tau\cdot b}t\right)\right]$	$b\cdot c+\tau$	$b^2\cdot c+\tau^2$	Exponentially modified gamma: this paper, see Section 2

* References: 1: Hawkins et al., 2007; 2: Duffy et al., 2012; 3: Dowling et al., 2014; 4: Maler and Lutscher, 2010; 5: Zilman et al., 2010; 6: Stukalin et al., 2013; 7: León et al., 2004; 8: Nakaoka and Inaba, 2014; 9: Leander et al., 2014; 10: Golubev, 2010a; 11: Tyson et al., 2012; 12: Gomes et al., 2011; 13: Castor, 1980.

assumed as evidencing the appropriateness of whatever cell cycle model that suggests a function showing such a fit. The gamma distribution is indeed derived from several published models whose justifications include satisfactory fits of resulting PDFs to experimental data. The Wald distribution (also known as reciprocal normal and inverse Gaussian) is used upon the assumption that the rates, not the times, of cell transit through the cell cycle are distributed normally. The lognormal distribution is adopted in several models of cell and tissue kinetics without much reasoning concerning underlying mechanisms but, rather, based on its assumed ubiquity in nature and on empirically found satisfactory fits to IDT data.

It has been shown that IDT data are usually approximated with the exponentially modified Gaussian (EMG) better than with other PDFs (Golubev, 2010a). EMG is consistent with the transition probability model of the cell cycle (TPM) (Smith and Martin, 1973). TPM implies that cells, on their way from birth by mother cell division to their own division, enter an intermediary state, somewhere within the G1 phase of the cell cycle, out of which they exit at random to the further succession of events culminating in mitosis. Because of the stochasticity of exiting the intermediate state, the time spent by cells there is distributed exponentially. This portion of the cell cycle constitutes the probabilistic phase A according to the original TPM. The rest of the cell cycle makes its deterministic phase B, which in the original TPM is assumed to feature a negligible variability. However, the time spent by cells in the phase B is variable, as all quantitative biological parameters are, and if this variability is high enough to be considered and, at the first approximation, is distributed normally, the entire IDT variability will be defined by the sum of two random variables whose distributions are exponential and Gaussian and, therefore, by the convolution of an exponential and a Gaussian distributions (Golubev, 2010a)].

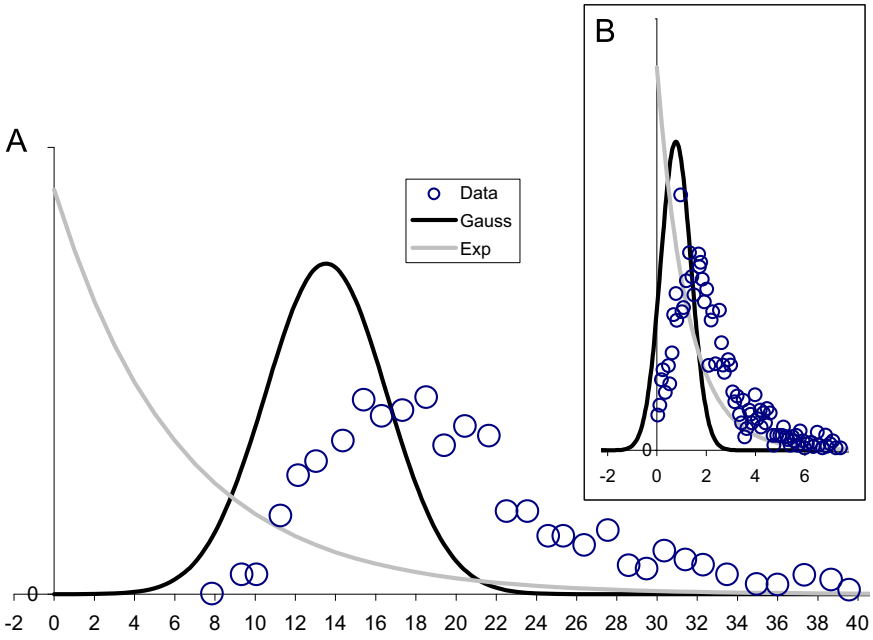

Fig. 2. Deconvolution of data on time distributions, which are assumed to correspond to EMG, into the Gaussian and exponential component. The meaning of data points is the same as in Fig. 1. A – Human B-cells IDT distribution (the same as in the lower right panel of Fig. 1): τ=6.15 h, S=13.53 h, σ=3.02 h. B – The distribution of times between transcriptional bursts of Bmal1 gene; data are derived from (Suter et al., 2011a): τ =1.24 h, S=0.76 h, σ=0.61 h.

Using EMG makes it possible to decompose an IDT distribution into the probabilistic and deterministic part of the cell cycle (Fig. 2A) and to distinguish the effects of, e.g., differentiation-inducing agents or anticancer drugs on the two parts (Golubev, 2010a; Tyson et al., 2012). The importance of such distinguishing follows from reasons to believe that increases in the duration of the A part rather than B part promote cell differentiation and/or cell population growth arrest (Golubev, 2010a).

After data have been published suggesting that the activity of some genes consists of stochastically variable alternating periods of engagement in transcription, which are manifested as transcriptional bursts, and of transcriptional idleness (Harper et al., 2011; Suter et al., 2011a, 2011b), and that the time distribution of the former is exponential and of the latter is peaked (Suter et al., 2011a, 2011b), it has been shown (Golubev, 2012a) that EMG may approximate idleness time distribution better than the convolution of two exponents or than the gamma distribution suggested by the authors of the original findings. A mechanistic rationale for EMG applicability to data on transcriptional discontinuity was suggested (Golubev, 2012a, 2012b). In this case, however, an inherent flaw of EMG becomes significant. A part of the Gaussian that emerges upon the deconvolution of an EMG extends to the negative time domain (Fig. 2b). This extension is negligible for IDT distributions whose peak positions (modes) are much greater than variances (Fig. 2a). Therefore, EMG may still be a practically acceptable approximation to IDT data.

However, strictly speaking, upon the assumption that an exponent is inherent in an IDT or other biologically relevant distribution, the exponent must be convoluted with a peak function whose domain is essentially nonnegative. In what follows, the appropriateness of the gamma distribution for this purpose will be explored and checked against a panel of published IDT data. Then the same approach will be applied to the analysis of time distributions related to gene activity. Finally, some implications of considering both, cell behavior and gene behavior, within this framework will be discussed.

2. Introducing the exponentially modified gamma distribution

The choice of the gamma distribution is supported by reasoning similar to that justifying some previously suggested models of intracellular processes (Floyd et al., 2010; Lee and Perelson, 2008; Schwabe et al., 2012; Weber et al., 2014). In essence, cell passage through the cell cycle may be envisioned as a series of c transitions through successive states, each transition being, at the cell population level, a first-order process constituted by random events; therefore, the probability density of the random time required to pass a transition is exponential:

$$\lim_{\Delta t\to 0}\left[\frac{N(t)-N(t-\Delta t)}{\Delta t}\right]=n(t)=\frac{1}{b}\cdot \mathrm{e}^{-\frac{t}{b}}$$

Assuming that the difference between any two of characteristic times b of the transitions is small compared with the smallest b, the probability density $n(t)$ for the time of going through all transitions may be approximated with the convolution of c exponential distributions having identical b. The result is known as the Erlang distribution:

$$n(t)=\frac{t^{c-1}e^{-\frac{t}{b}}}{b^c(c-1)!}$$

This is a particular case of the gamma distribution having only integer values of its shape parameter c:

$$n(t)=\frac{t^{c-1}e^{-\frac{t}{b}}}{b^c\Gamma(c)},$$

where $\Gamma(c)=\int_0^\infty t^{c-1}e^{-t}\mathrm{d}t$

If the characteristic time (CT) of one of the transitions is so long that the difference between the CTs of any two of other transitions may be neglected, compared with the difference between the CT of the slowest and of the next slow transition, and is independent from other CTs, then the overall IDT distribution may be approximated by the convolution of the gamma distribution capturing the relatively rapid transitions and of the exponent capturing the slow transition. An illustration of the above is presented in Fig. 3, which shows the convolution $L(t)$ of an exponential

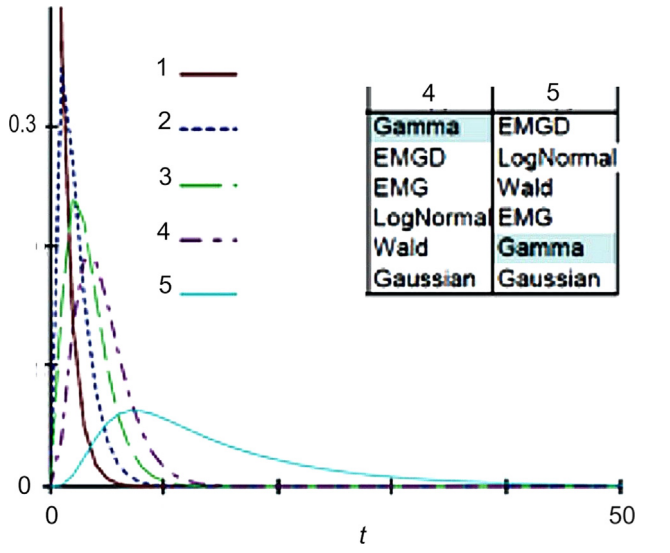

Fig. 3. Sequential convolution of five exponential distributions: 1)$H(t)=e^{-t}$; 2)$I(t)=H(t)*\frac{1}{1.1}e^{\frac{-t}{1.1}}$; 3)$J(t)=I(t)*\frac{1}{1.2}e^{\frac{-t}{1.2}}$; 4)$K(t)=J(t)*\frac{1}{1.3}e^{\frac{-t}{1.3}}$; and 5)$L(t)=K(t)*\frac{1}{10}e^{\frac{-t}{10}}$. The characteristic times of four of the exponents convoluted in $H(t)$ to $K(t)$ are similar, whereas that of the exponent convolved with $K(t)$ to get $L(t)$ is much longer. The inset shows the results of approximating $K(t)$ and $L(t)$ with several peak function arranged top to bottom according to their fits. See the main test for more explanations.

distribution whose CT is 10 with the convolution $K(t)$ of four exponential distributions whose CTs are 1.0, 1.1, 1.2, and 1.3.

Both $L(t)$ and $K(t)$ are, strictly speaking, hypoexponential distributions. Closed-form solutions for some particular cases of the hypoexponential distribution have been suggested (Weber et al., 2014). However, concerning the cell cycle, there is no *a priori* idea about the number of exponents to be convolved and about the relationships between their parameters. It was assumed in (Weber et al., 2014) that shifted exponential distributions characterize the variabilities of times of cell transits through G1, S and G2+M, leading to a shifted hypoexponential function, which results from the convolution of three exponents. Here, the decision to lump G2 and M together is questionable.

A milder assumption that the CT of one of the convolved exponents is much greater compared with all others provides for a more flexible approach to fitting. Indeed, a gamma distribution is distinguished as the best-fit function for $K(t)$. The estimates of its b and c are 1.15 and 3.99, respectively, the latter figure corresponding to the number of convoluted exponents, and the former, to the mean of their CTs. The estimates of EMGD parameters (see below) for $K(t)$ are $\tau=b=1.14$ and $c=3.03$, that is, three exponents are convoluted with a similar one, which is also in accordance with the input assumptions. $L(t)$ is best fitted with an EMGD having $\tau=9.96$, $b=1.17$, and $c=3.91$. The gamma distribution approximates $L(t)$ worse, the estimates of b and c being 4.47 and 2.84, respectively.

Under the above assumptions, the convolution integral for $n(t)$ will be:

$$n(t)=f(t)*g(t)=\int_0^t f(t-x)\cdot g(x)\cdot \mathrm{d}x=\int_0^t \frac{\mathrm{e}^{-\frac{t-x}{\tau}}}{\tau}\cdot\frac{x^{c-1}\mathrm{e}^{-\frac{x}{b}}}{b^c\Gamma(c)}\cdot \mathrm{d}x,$$

where $f(t)$ is the exponential distribution, which captures the slowest transition, and $g(t)$ is the gamma distribution, which captures c fast transitions.

The solution of the above improper integral may be written as:

$$n(t)=e^{-\frac{t}{\tau}}\cdot b^{-c}\cdot\left(\frac{\tau-b}{\tau\cdot b}\right)^{-c}\cdot[\tau\cdot\Gamma(c)]^{-1}\cdot\left[\Gamma(c)-\Gamma\left(c,\frac{\tau-b}{\tau\cdot b}t\right)\right] \quad (1)$$

where $\Gamma(c,\frac{\tau-b}{\tau\cdot b}t)=\int_{\frac{\tau-b}{\tau\cdot b}\cdot t}^{\infty} e^{-t}\cdot t^{c-1}\cdot \mathrm{d}t$ is an upper (complementary) incomplete gamma function.

Algorithms for numerical treatment of incomplete gamma functions are included in standard curve-fitting software tools. To reduce the number of adjustable parameters of the incomplete gamma function, Eq. (1) is reparameterized as:

$$n(t)=e^{\frac{-t}{\tau}}\cdot\left(\frac{\tau\cdot B+1}{\tau\cdot B}\right)^{c}\cdot[\tau\cdot\Gamma(c)]^{-1}\cdot[\Gamma(c)-\Gamma(c,B\cdot t)] \quad (2)$$

where $B=\frac{\tau-b}{\tau\cdot b}\cdot$

3. Treating published interdivision time data with the exponentially modified gamma distribution

3.1. Preliminary considerations

Making inferences about the nature of processes underlying an observed distribution of variable parameters from the shape of such distribution is a routine approach in physics. In fact, the software tools designed to judge what functions are more appropriate to a given dataset were developed for physical applications, as exemplified with the tool used in the present study: TableCurve 2D (TC2D) ver. 5.1 (Systat Software Inc., San Jose, CA, trial version is available at http://www.sigmaplot.com/products/tablecurve2d), which allows comparing its more than 40 inbuilt peak-functions, including EMG, with user-defined functions, such as Eq. (2). Nonlinear maximum likelihood estimation afforded by TC2D was executed using the least squares minimization option.

In a sense, peering into the details of distributions and discriminating between them using a curve-fitting software is similar to peering into the details of cell structures with an electron microscope. Even users who are superficially familiar with the principles of electron optics can draw reasonable conclusions from such observations. Instead of checking all instrument settings and underlying physical details in order to be sure that an observed feature is not an artifact, one can employ appropriate controls. Indeed, such controls are required in either case. For example, when using any software, it is reasonable to check its ability to discriminate different artificial distributions deliberately generated by numerical modeling. In particular, Eq. (2) introduced as a user-defined function in TC2D software was found to exactly fit artificial data generated using Eq. (1) and to return the estimates of τ, b and c values that were equal to the respective input values. At the same time, Eq. (2) was found to be inferior to gamma, lognormal, EMG, Wald or any other distribution by its fits to data generated with the respective functions.

This a sort of computer-assisted experimental mathematics as discussed in Avigad (2008) is now increasingly applied in very different biological contexts, such as molecular cell biology (Geiler-Samerotte et al., 2013) and psychophysiology (Palmer et al., 2011). In particular, convolutions of exponent with various peak functions are used to describe, respectively, the distributions of times of cell birth to cell division (interdivision times, IDT) (Golubev, 2010a; Tyson et al., 2012; Leander et al., 2014) and of times from stimulus to response (response times, RT) (Palmer et al., 2011). Arguments concerning the use of this approach in the two fields are remarkably similar:

"How impressed should one be by the fits of these functions to the data and by the differences between the fits of the functions? Can this approach be used to reject hypotheses about the RT distributions? … The ex-Gaussian function yielded the closest fit to the data, overall, followed by the ex-Wald, Gamma, and then the Weibull function, though there is little difference between the performances of the first three functions. The close ties of the ex-Wald distribution to diffusion process models makes this function

particularly relevant to search, with a clear set of interpretations for its parameters… The ex-Gaussian function also fits these and other RT distributions well…, and the theoretical status of the individual parameters in the ex-Gaussian has been well-explored by many researchers … However, the ex-Gaussian parameters are not as strongly tied to computational models of search as the ex-Wald parameters" (Palmer et al., 2011).

Questions like those cited above will be addressed in Sections 3.2 through 3.4 with regard to the applicability of the exponentially modified gamma distribution (Eq. (1)) compared to other functions (Table 1) used to treat data on the variability of times from cell birth to cell division, and in Section 4, with regard to distributions of times between the busts of mRNA production by active genes.

3.2. General analysis

Interdivision time (IDT) data are virtually always presented as histograms showing the numbers or proportions of cells falling into specified IDT intervals. Since mid-1960s, when time-lapse cinematography was introduced for individual cell tracking, such data on eukaryotic cells have been published by 10 or so research groups in about a score of papers reporting on results obtained with appropriately high numbers of cells. Altogether, the evidence analyzed here (see supplementary Table S1) comprises 77 datasets related to 16 cell types, some studied under different conditions.

These datasets were digitized by taking the midpoints of the tops of histogram bars. It was assumed that zero is the observed number of cells in the time intervals immediately before and after the first and the last interval where at least one cell is present and that there are no observations further on to the left and to the right of a histogram. The resulting data were fit with EMGD as a user-defined function (Eq. (2)) introduced into the TC2D software package and compared with five of 40 peak functions inbuilt in TC2D. The functions chosen for comparison have already been used by different authors to treat IDT data (see Table 1): simple Gaussian, EMG, and lognormal, gamma and Weibull distributions, the latter two being available in TC2D in their shifted forms (τ–x instead of τ as the argument, where x is another adjustable parameter). This panel was supplemented with two more user-defined functions: pure (non-shifted) gamma distribution and Wald distribution, which were used earlier by other authors for fitting IDT distributions (see Table 1).

Two goodness-of-fit criteria available in TC2D were used: determination coefficient unadjusted for the degrees of freedom (r^2=1–SSE/SSM) and f-statistic (F=((SSM–SSE)/(m–1))/(SSE/DOF)), which accounts of the degrees of freedom and the number of adjustable parameter (DOF=n–m, where n is the number of data values and m is the number of adjustable parameters) and thus gives preference to functions having fewer parameters to be adjusted. For every dataset analyzed, the eight functions were ranked according to their fits and ascribed with numbers from one (the best fit) through eight (the worst fit).

The overall comparison of the conformances of the functions tested against the whole body of available evidence is presented in Fig. 4.

According to the determination coefficient, either EMG or EMGD appears to be the function of choice as a single model for all datasets.

Being suggestive in this regard, Fig. 4 however provides no cues concerning the statistical significance of what it suggests. In almost all cases the differences between the fits of any two of the eight functions to a specific dataset are not significant whatever parametric criterion is used. Moreover, different datasets comprise different numbers of data points and relate to experimental settings associated with different levels of data noisiness; therefore, the ranges of parametric fit criteria accommodating the estimates of fits of all functions tested vary markedly between different datasets. That is why the issue of superiority of one function over another was addressed using ranking, and the issue of whether superiority is not accidental was tackled using nonparametric statistics. Because the number of datasets studied is high enough, even the simplest nonparametric criterion, i.e. the sign test, may be employed for that. Each function in a list ranked according to the fits of the functions to a dataset was compared with each of the other functions. The sign "minus" was assigned to a function standing lower in a list than a function taken for comparison. With 77 datasets, the null hypothesis (the superiority of a given function

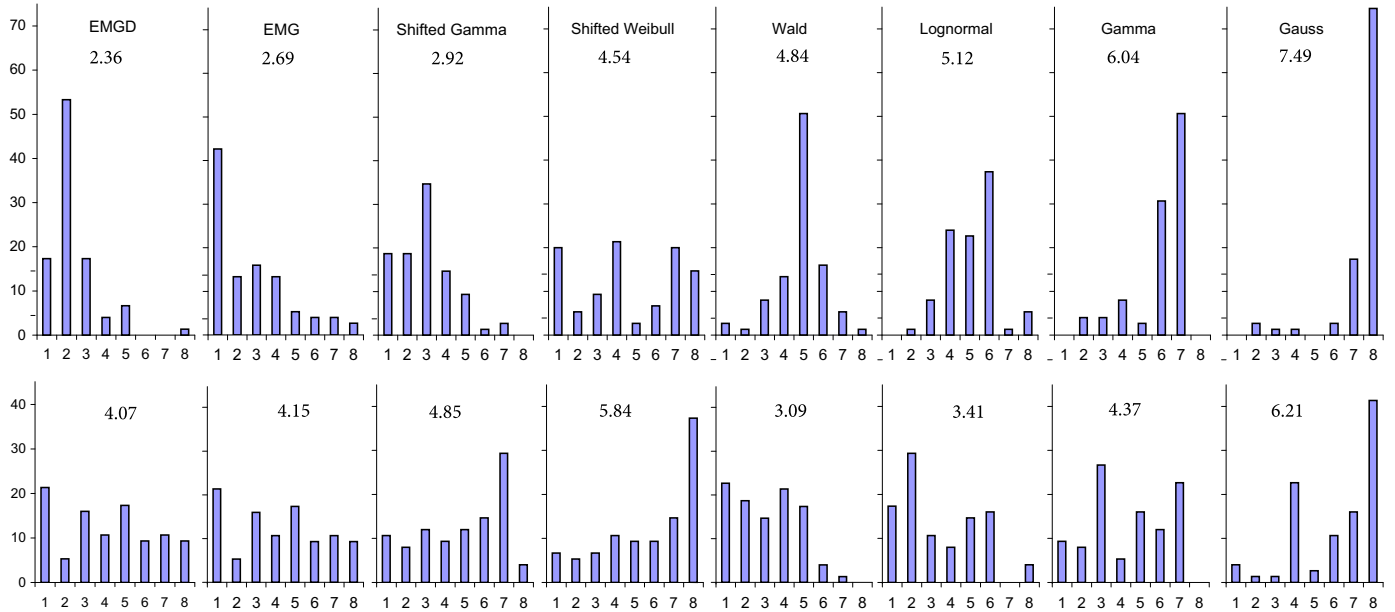

Fig. 4. Integral evidence of the conformance of eight functions to seventy-seven IDT distributions. Vertical bars show the proportions (percent) of IDT datasets whose ranks of fit to a given function are indicated at bar bottoms. The mean rank of a function is shown above the respective fit distribution. The upper and the lower row show fits according to the determination coefficient (r^2) and f-statistic, respectively. Raw data are presented in Supplementary Table S1.

Table 2
Employing the sign test for pairwise comparisons of the performances of functions used to approximate IDT distributions.

	r^2								*f*-statistic						
	EMG	s-Gamma	s-Weibull	Wald	Lognorm	Gamma	Gauss		Lognorm	EMGD	EMG	Gamma	s-Gamma	s-Weibull	Gauss
EMGD	*37*	**22**	**20**	**7**	**5**	**5**	**3**	Wald	*35*	*32*	*34*	**13**	**25**	**19**	**4**
EMG		***27***	**21**	**11**	**10**	**9**	**2**	Lognorm		*34*	*36*	**17**	**29**	**18**	**8**
s-Gamma			**22**	**8**	**8**	**7**	**5**	EMGD			*35*	*40*	**26**	**19**	**24**
s-Weibull				*34*	*31*	***27***	**13**	EMG				*38*	*30*	**21**	**25**
Wald					*35*	***11***	**4**	Gamma					*30*	**23**	**3**
Lognorm						***16***	**8**	s-Gamma						**23**	***29***
Gamma							**3**	s-Weibull							*37*

The functions are listed in the order of their increasing mean ranks according to r^2 (left panel) or *f*-statistic (right panel). A function shown in the left column performs significantly better than a function shown in the upper row if the respective number is <26 ($p<0.01$, bold) or <29 ($p<0.05$, bold italic). Insignificant differences are shown with thin italic.

is accidental) is rejected at $p<0.05$ or $p<0.01$ when the number of minuses is less than 29 or 26, respectively (Dixon and Mood, 1946) (Table 2).

According to r^2, EMGD and EMG are equivalent and outperform the nearest competitor at $p<0.01$. According to *f*-statistic, it may seem, albeit less definitely, that the two functions having fewer parameters, Wald and lognormal, may be more preferable based on greater skewenesses of their fit rank distributions towards higher ranks and on their higher mean ranks (Fig. 4). However, their superiority is insignificant when assessed with the sign test. Apart of formal criteria, the assumption that fewer parameters are better becomes less sensible when there are more reasons to believe that an additional parameter captures some essential feature of the phenomenon in question. Therefore, preference is given below to the determination coefficient.

Noteworthy is that EMG often stands higher than EMGD when both are found at the two highest rank positions. This may be because numerous confounding effects can produce a sort of Gaussian blur of the peak component convolved with the exponential component of an entire IDT distribution and thus can make the shape of the peak component, which otherwise would better conform to a gamma distribution, be closer to a Gaussian, and the whole IDT distribution be closer to an EMG.

Using EMGD, so as any other function, to approximate distributions that tend to take an EMG-like shape may lead to potentially misleading observations. A likely case is illustrated in Table 3, which shows the results of numerical experiments where EMGD was used to approximate EMG having $S=5$, $\sigma=1$, and τ varying from 1 to 10. The artifacts of such approximation include τ underestimation, which is most apparent at low τ/S, a negative correlation between τ and b, and a positive correlation between τ and c.

Another likely cause of systematic deviations of real IDT distributions from whatever model used to approximate them is that many of the available data relate to cell populations comprising several cell generations. Even if each generation exactly conforms to an EMGD (or another function), this does not necessarily mean that the parameters of the EMGD (or the other function) are the same for all generations. In fact, in a few studies where defined cell generations were examined, trends to either decreases or increases in the mean IDT of successive cell generations may be noticed (see Section 3.2 below). As far as EMGD is concerned, the number of processes involved in the "deterministic" part B of the cell cycle is likely to be the same in a defined cell type; therefore, changes in the mean IDT may be attributed not to c, but to τ or/and b.

Numerical modeling shows that when an artificial distribution that represents a mixture of EMGDs having variable τ and constant b and c values is presented to a curve-fitting software, the fit of an EMGD to such distribution, although not quite exact, will still be the best compared with other functions, among which an EMG

Table 3
EMGD as an approximation of EMG having $S=5$, $\sigma=1$, and variable τ.

τ_{EMG}	**EMGD parameters**				
	τ_{EMGD}	***b***	***c***	***S***	σ^2
1	0.653	0.246	21.63	5.329	1.313
1.2	0.961	0.233	22.36	5.218	1.218
1.5	1.335	0.221	23.28	5.152	1.140
2	1.866	0.216	23.70	5.116	1.104
4	3.916	0.208	24.33	5.063	1.053
6	5.931	0.203	24.86	5.058	1.029
10	9.909	0.202	24.95	5.049	1.022

will be the next best-fit function. In this case, the value of τ will be within the range of τ values featured by the component distributions comprising the mixture, whereas b will be somewhat higher and c will be somewhat lower than the respective parameters of the component distributions. For example, the mixture of four EMGDs whose b is 1, c is 20, and τ are 3, 5, 7, and 10, will be recognized as an EMGD having $b=0.892$, $c=21.691$, and $\tau=6.610$. When a mixture of EMGDs differing in their b is examined, its fit to EMGD will be inferior to its fits to gamma, lognormal and Wald distributions, and EMG will show the worst fit, except for that shown by a pure Gaussian. In this case, the yielded τ will be markedly higher than the initial τ; b will be also higher, outside the input range; and c will be lower. For example, at the initial $\tau=10$, $c=20$, and b being 1, 1.2, 1.5 and 2.0, EMGD fitted to the mixture of distributions will have $\tau=13.328$, $c=9.765$, and $b=2638$.

With the above caveats in mind, it is unreasonable to expect much from inferences about b and c values and their changes extracted from IDT data. Only τ of the exponent component of EMGD and the mean (S) of the gamma function convoluted with it, that is, the mean durations of the probabilistic and deterministic part of the cell cycle may be relied on. In this regard, EMGD and EMG are virtually equivalent as descriptive tools for cell cycle time variabilities.

3.3. Specific applications

The usability of EMGD for analyzing IDT results obtained with a defined cell type under different conditions will be examined below starting with data judged as more reliable for the following reasons:

- higher numbers of more homogenous and robustly proliferating cells must reduce noise;
- longer observation times must reduce the bias resulting from the exclusion of longer IDTs; and

Table 4
HeLa cells IDT data derived from histograms published in (Hurwitz and Tolmach, 1969).

Cell condition	Cell generation	Cells number	Fit rank by r^2								Parameters of EMGD (h, except for c)					
			EMGD	EMG	Wald	s-Gamma	Gamma	Lognorm	s-Weibull	Gauss	τ	b	c	S	σ	IDT
Control cells	**1**	*342*	3	4	5	2	7	6	1	8	2.311	0.094	176	16.749	1.261	19.06
	2	*320*	2	1	5	4	6	8	3	7	1.512	0.078	212	16.473	1.131	17.985
	3	*132*	2	1	5	4	7	3	6	8	0.754	0.11	153	16.882	1.364	17.636
	1+2+3	*794*	2	3	4	1	6	5	7	8	2.048	0.075	218	16.499	1.116	18.547
Irradiated cells	**1**	*253*	3	4	5	1	7	6	2	8	2.407	0.226	72	16.243	1.916	18,65
	2	*194*	2	1	5	3	6	4	7	8	3.734	0.255	69	17.63	2.12	21,364
	3	*98*	2	3	4	1	6	5	8	7	3.954	0.211	76	16.107	1.844	20,061
	1+2+3	*545*	1	3	5	2	7	6	4	8	3.111	0.237	71	16.779	1.995	19,89
	All before division	*599*	1	3	5	2	7	6	4	8	3.388	0.234	70	16.376	1.956	19.764
	All before death	*229*	1	3	5	2	7	6	4	8	5.878	0.547	30	16.609	3.015	22.487
Mean fit rank			1.9	2.6	4.8	2.2	6.6	5.5	4.6	7.8						

– higher fit ranks of EMGD and/or EMG compared with alternative models must provide for a more trustworthy estimation of changes in the probabilistic and deterministic part of the cell cycle.

In this regard, the study reported as early as by the end of 1960s (Hurwitz and Tolmach, 1969) looks as if deliberately designed to evaluate the applicability of different models to IDT distributions. HeLa cells used in the study are generally believed to feature a remarkably robust proliferation. High numbers of cells tracked using time-lapsed cinematography made it possible to attribute appropriately high numbers of IDTs to each of three cell generations. Although the times between successive premitotic cell roundings rather than divisions were measured in this study, this is unlikely to compromise inferences from averaged data significantly.

Data on the fits of different models to HeLa cells IDT distributions are presented in Table 4. Based on the results related to any single generation (F), one may choose either shifted Weibull (control F1), or EMG (control F2 and F3), or shifted gamma distribution (irradiated F1 and F3) to treat the data. However, upon the assumption that an IDT distribution is shaped by the mechanism of the cell cycle, it seems unlikely that each generation of robustly proliferating HeLa cells has a mechanism of its own. If a common model for all generation has to be chosen, the reasons presented above suggest the EMGD for this purpose.

Using EMGD reveals a consistent trend to decreasing τ in successive generations of control cells and the opposite trend in irradiated cells, there being no consistent changes in S and in the entire IDT. Because only τ is influenced consistently, pooling of data related to different F will not considerably violate EMGD (see above 3.2). Such pooling shows that irradiation increases τ by 50% without influencing S. In the irradiated cells that die after premitotic rounding, τ is by 73% higher than in cells that go on to divide, there being no difference in S. This sensitivity of τ but not of S to changes in cell conditions is reproduced when EMG is used for data treatment (not shown) and is obscured with other functions.

The apparent negative correlation between b and c is most likely caused not by the cells as they are, but by using EMGD to treat data biased toward EMG. Therefore, the only reliable inference from b and c estimates is that the intracellular events contributing to the deterministic phase B of the cell cycle are numerous indeed and that the processes comprised of these events in cell populations are 10–20 times faster compared with the process responsible for the probabilistic part A captured by τ.

Table 5
Stimulated B-cells IDT data derived from histograms published in *Duffy et al. (2012) and **Hawkins et al. (2009).

Stimulus	Cells generation or #condition		EMGD parameter (h)			
			τ	S	σ	IDT
*CD40 ligand	1		9.565	17.313	4.766	26.878
	3		5.885	13.816	3.113	19.701
	5		3.674	10.841	1.150	14.515
	7		6.073	8.818	0.571	14.891
	All		10.224	10.623	1.644	20.847
**CpG	1		4.658	33.38	1.89	38.038
	All-1		3.423	6.59	0.94	10.013
	#Daughters divided:	two	1.295	7.51	1.068	8.805
		one	2.307	8.42	0.49	10.727
		zero	2.166	9.54	1.949	11.706
		two	1.971	6.903	0.209	8.874
		one	3.133	7.949	0.726	11.082
		zero	2.992	9.654	1.972	12.646
	#Sisters divided:	both	4.033	6.869	0.75	10.902
		one	5.088	10.846	1.42	15.934
		both	4.386	6.864	0.17	11.25
		one	7.133	11.954	0.256	19.087

The estimates of c are high enough to make the Gaussian a good approximation of the gamma distribution, since it is generally believed that at $c > 30$ it is virtually indistinguishable from the normal distribution – another reason to consider EMG a practically good model of IDT distributions. In fact, the estimates of S and τ obtained with EMGD and EMG are almost identical (Supplementary Table S1).

The artifactual negative correlation between b and c estimates tempers the temptation to interpret the apparent differences in b and c between control and irradiated cells as an evidence that, in the latter, fewer processes constitute the deterministic part B of the cell cycle. Since similar reasoning is applicable to all data considered below, only S and σ (the artificial correlation between c and b is partially annihilated by lumping b and c together in these parameters) will be considered to avoid encumbering.

One more source of IDT data related to several defined generations of cells is the study of B-lymphocytes stimulated with a CD40 ligand (Duffy et al., 2012). Using EMGD for the same purposes and reasons as above, and with the same caveats, shows that, this time, the deterministic part B rather than the probabilistic part A of the cell cycle decreases in successive cell generations (see Table 5).

Since the clonal expansion of B-cells is a programmed process realized within a limited number of cell divisions, this decrease

may be a part of the program. Among the successive generations, F1 stands out in having markedly high S and τ, probably because F1 IDTs include an additional randomly varying periods required for stimulated cell to enter the cell cycle. As a result, the pooled distribution, although it conforms to EMGD, features an exceptionally high τ, outside the range of τ values of the component distributions, including even F1. This should be expected based on numerical experiments discussed in 3.2, which also suggest that, in this case, pooling of data related to several generations having different b may be misleading, especially when F1 data are included. In another publication by the same group (Hawkins et al., 2009) pooled IDT data for CpG-stimulated B-cells do not include F1 cells, which are presented separately. In this case, too, F1 cells have markedly high τ and S, and this feature is reproduced with both, EMGD and EMG. However, in both papers, IDT data were fitted with the lognormal distribution, based solely on its good fit to data, which is actually better only in comparison with the pure Gaussian and is clearly inferior to EMGD, EMG, Wald and shifted gamma distribution.

In the paper (Holzel et al., 2001) where the effects of c-Myc and Mad proteins on cell proliferation were reported, IDT distributions were presented as histograms where each bar showed the cumulative number of cells ascribed to several successive generations. Data on each generation derived from the histograms by subtraction are shown in Table 6, where the term “all generations” refers not to the sums of F1 and F2 but to the tops of the bars additionally comprising the summed cells referred to generations F3 and higher, which are not numerous enough to be treated separately.

Using EMGD (Table 6) and EMG (Golubev, 2010a) to deconvolve IDT distributions into the probabilistic and deterministic part of the cell cycle shows that c-Myc knocking-out results in a marked increase in τ, whereas S remains almost constant. This effect of attenuated Myc expression is confirmed by analyzing IDT of cells were *c-myc* gene was modified to make its expression be sensitive to tamoxifen. The antagonistic effect of Mad on Myc functions is manifested as an increase in τ. There is no evidence of an increase in S in F1 cells under the effect of Mad expression; however, pooling up of all generations suggests that such increase is possible, although such procedure is likely to be prone with artifacts.

An opportunity to make inferences about drug effects on cancer cells from changes in the shapes of IDT distributions is provided by data summed up in Table 7.

As suggested by Table 7, Erlot levels that produce no increase in S can cause 2- to 3-fold increases in τ, whereas cycloheximide (CHX) increases S more than τ. Thus, Erlot influences primarily the probabilistic part A of the cell cycle, whereas CHX acts at the deterministic part B. The same conclusions were drawn based on EMG (Tyson et al., 2012). Raltitrexed effect on human ileocecal carcinoma cells IDT is fully accounted for by a more than twofold increase in τ with no contribution from changes in S.

It may be seen in Tables 4–7 that τ is more sensitive than S to changes in cell conditions and that even a small relative increase in the mean IDT may be associated with a several-times greater relative increase in τ, if τ, not S, is responsible for changes in IDT. It was reasoned (Golubev, 2010a, 2010b) that changes in τ rather than in S can alter cell propensity for differentiation and/or other fate options competing with proliferation. Indeed, the ability of Erlot to potentiate the effects of differentiation inducers has been reported (Lainey et al., 2013). As to raltitrexed, it was noted (Slocum et al., 2000) that this drug not only increased IDT (by increasing τ, see Table 7) but also increased the proportion of cell that did not divide yet stayed viable and acquired distinctive morphological characteristics – the features generally associated with cell differentiation or senescence.

3.4. Implications for cell cycle structure

On a whole and with all the above caveats, the two exponentially modified peak functions, EMGD and EMG, included in the panel of eight functions compared in the present study, significantly outperform any other function in the ability to approximate IDT distribution (Table 2) and appear to be the best or second best-fit functions for IDT distributions in about 2/3 of all

Table 7
Anticancer drugs effects on EMGD parameters of IDT distributions.

Cell type and data source	Treatment	EMGD parameter (h)			
		τ	S	σ	IDT
Transformed human mammary epithelial cells(Tyson et al., 2012)	DMSO	2.920	12.936	0.841	15.856
	1 μM Erlot	3.829	13.304	0.956	17.133
	8 μM Erlot	9.992	16.666	2.913	26.658
	16 μM Erlot	8.640	25.625	2.064	34.265
Human lung adenocarcinoma (Tyson et al., 2012)	DMSO	6.214	19.4	2.352	25.614
	15.6 nM Erlot	12.487	17.634	1.864	30.121
	62.5 nM Erlot	10.034	22.277	5.391	32.311
Human mammary epithelium(Tyson et al., 2012)	DMSO	0.821	7.705	1.409	8.526
	Erlot	4.762	6.960	0.895	11.722
	CHX	2.632	15.000	2.400	17.632
Unspecified cancer(Leander et al., 2014)	DMSO	0.947	11.914	0.959	12.861
	Erlot	3.170	11.984	0.984	15.154
	CHX	4.037	20.487	3.453	24.524
Ileocecal carcinoma(Slocum et al., 2000)	Control	2.115	8.681	1.112	10.796
	Raltitrexed	4.749	7.985	1.517	12.734

Table 6
IDT data for cells differing in the conditions of their Myc and Mad proteins. The data are derived from histograms published in Holzel et al. (2001).

Cell type	Cell condition	Cell generation	EMGD parameter (h)			
			τ	S	σ	Mean IDT
Rat fibroblasts	Myc^{++}	1	5.937	20.922	2.794	26.859
		All	5.869	21.119	3.031	26.988
	Myc^{--} by knocking out	1	10.793	22.517	1.25	33.310
		All	10.114	22.894	1.24	33.008
	Myc^{+} by upregulation	1	5.437	19.619	2.889	25.056
		2	2.613	22.07	3.788	24.683
		All	3.665	20.947	3.416	24.612
	Myc^{-}	1	12.298	18.317	1.022	30.615
		2	10.074	20.827	3.506	30.901
		All	9.995	17.939	2.135	27.934
Human osteosarcoma	Mad^{-}	1	3.443	21.676	2.167	25.119
		All	3.887	20.487	1.96	24.374
	Mad^{+} by upregulation	1	11.248	22.95	1.82	34.198
		All	5.381	25.907	2.947	31.288

cases examined. This makes a reason to regard the exponent decay function as the immanent component of any IDT distribution, although it may be obscured by the effects of numerous confounding factors. Therefore, a relatively long-dwelled state exited by cells at random, which is captured by the exponent, must be a significant component of the cell cycle.

Grounds for regarding this state as associated with the restriction point (RP) have been provided (Golubev, 2010a, 2010b, 2012a, 2012b), with the caveat that the moments of cells transitions to being committed to the S phase or alternative options are not confined to a certain position in G1. Only the time when such transitions become possible, after a cell has been born by division of its mother cell, can be confined (with reservations accounting of the variability of the preceding phase). Therefore, the conventional term "restriction point" refers to what in essence is rather a "fate-defining state" whose duration is variable and, in a single cell, may be determined only post-factum. In a cell population, its duration is distributed exponentially.

If RP in the above sense is still mapped to G1, then G1 time distribution must conform to an exponentially modified peak function, such as EMGD, whereas times distributions of other phases must not be skewed thus heavily.

In the literature, controversies concerning RP range from debating RP position in G1, the number of RPs at work under specific conditions, and RP relationships with other cell cycle checkpoints (Cooper 1982; Lee and Perelson, 2008) to questioning the mere existence of an RP (Cooper, 1982, 2015).

Different views imply different distributions of the times of separate phases of the cell cycle or of their combinations (see Table 1). In this regard, it is noteworthy that the time distribution of the whole cell cycle must be defined by the convolution of the functions chosen for time distributions of its separate phases. Such convolution will never yield the same function as any of the functions convolved, if only not to admit that all phases feature the Gaussian distribution. Finding a closed-form formula for a convolution is not a trivial task, especially if there are more than two functions to convolve, and even more so, if they are not all the same. However, to use a "unitary" function for IDT distribution means to imply that the cell cycle does not consist of discernible consecutive phases (stages, parts).

The same holds true for any of cell cycle phases. If G1 includes a step passed in a way different from transiting the rest of G1, then G1 time distribution is unlikely to conform to a "unitary" distribution, such as inverse normal (Castor, 1980; Cooper, 1982; Leander et al., 2014) or lognormal (Dowling et al., 2014).

The former, also known as Wald distribution, is implied in the "continuum model" of G1 (Cooper, 1982, 2015) based on the premise that the rates, not times, of uninterrupted transit through G1 are distributed normally.

The usability of the latter, which is included in the "stretched model" of the cell cycle (Dowling et al., 2014), is supported only by curve fitting. The model assumes that the time variabilities of any phase of the cell cycle and of the whole cell cycle conform to the same type of distribution, which is stretched proportionally to the

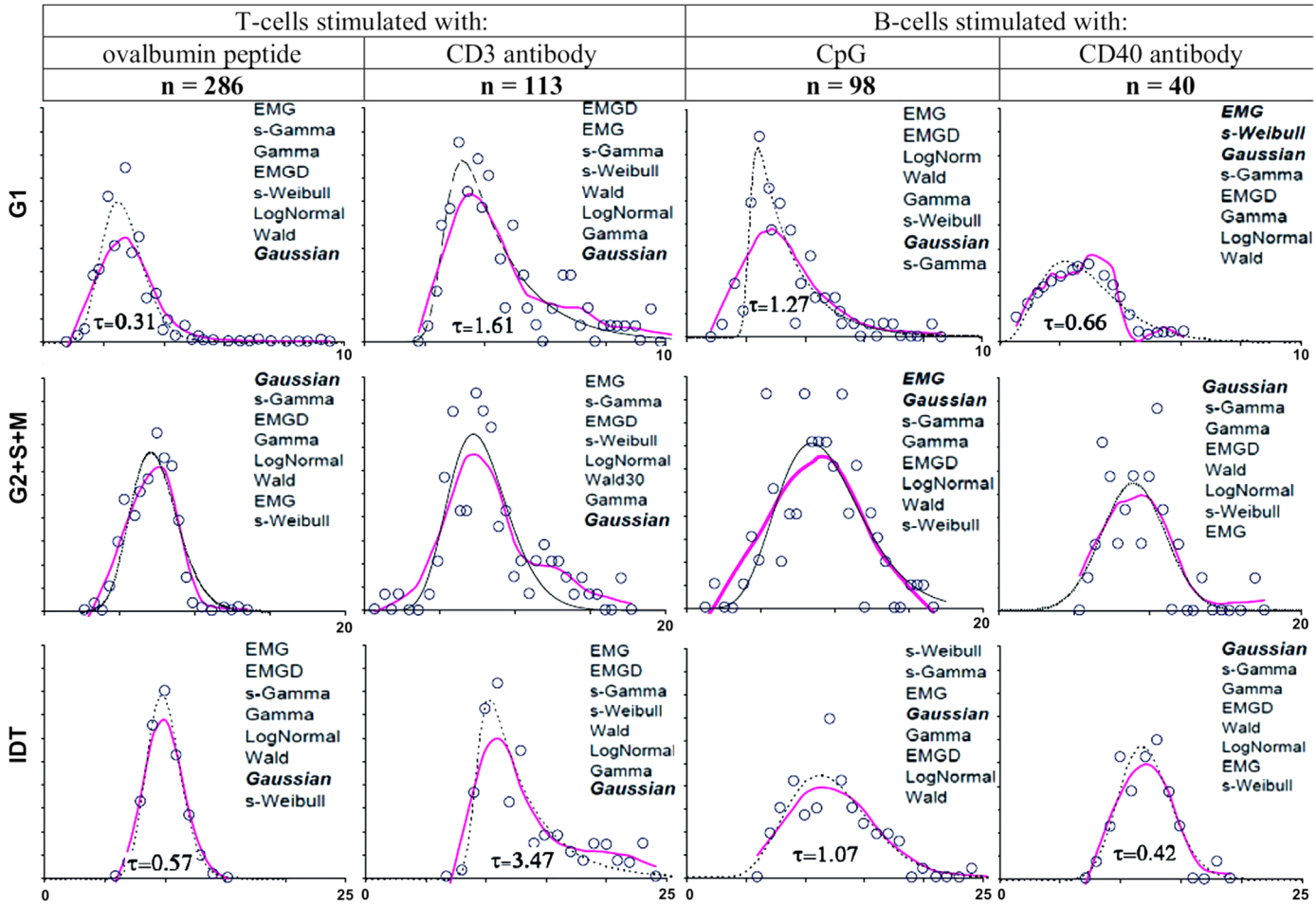

Fig. 5. Time distributions of defined phases of the cell cycle. *Open circles*: data points corresponding to the tops of the bars of histograms presented in Fig. S6 of (Dowling et al., 2014). *Thick solid lines*: LOESS smoothing of data. *Thin/dotted lines*: approximations of data with EMGD (for G1 and IDT) or gamma distribution (for G2+S+M). *Inserts*: functions tested for their ability to approximate data are arranged according to their fits to data. Bold italic characters designate symmetric and left-skewed functions (TC2D can approximates both left-skewed and right-skewed functions with EMG, the former having negative values of τ).

mean duration of each phase and of their sum. Direct data on G1 timing were obtained by experiments with lymphocytes gene-engineered to express a fluorescent protein in the period starting from the onset of the S-phase and continuing over the rest of the cell cycle (S+G2+M). This allowed distinguishing the G1 phase of individual cells by subtracting the S/G2/M time from IDT determined by direct cell visualization.

Time distribution data presented in Dowling et al. (2014) with bar histograms are presented here in Fig. 5 as open circles against time scales, which for G1 and G2+S+M data are stretched to align them with those for IDT data, in order to facilitate comparing the shapes of all distributions. To obtain a possibly less preconceived preliminary idea about the skewnesses of the distributions prior to fitting them with any model, they are smoothed by applying the locally estimated scatterplot smoothing procedure (LOESS) implemented in the TC2D software tool. The result presented as thick solid lines makes the impression that G1 and IDT distributions are more right-skewed compared with S/G2/M distributions, which is confirmed by comparing the fits of the data to defined distributions. Right-skewed distributions, including EMG and EMGD, are found at the tops of ranked lists related to G1 and IDT data. An exception is CD40-stimulated B-cells, likely because cells number ($n=40$) is too small to make apparent time distributions accurate. As to S/G2/M, the best-fit EMG and s-Weibull distributions found at the tops of ranked lists are, in these cases, actually left-skewed (the estimates of τ for EMG are negative).

Taken together, data treated in the above way do not support the idea that a single type of time distributions is appropriate to all phases of the cell cycle and to the whole IDT. Moreover, even if such assumption were acceptable with regard to separate phases, the whole IDT distribution would conform not to the chosen distribution but to the convolution of two or more of such distributions, depending on the assumed partitioning of the cell cycle. If it is assumed that G1 and S/G2/M time distributions are lognormal, as it is done in (Dowling et al., 2014), it may be easily shown by numerical modeling that their convolution (IDT distribution) will be best approximated with a shifted gamma distribution.

The results of treating data from (Dowling et al., 2014) according to TPM are presented in Fig. 5 with thin/dotted lines. Because G1 consists of the part A and a portion of the part B of the cell cycle, G1 data must be fitted with an EMGD, not with an exponential distribution, which is used in (Dowling et al., 2014) to treat G1 data as if according to TPM. S/G2/M data are fitted with the gamma distribution. IDT data are fitted not with the convolution of EMGD and gamma distribution related to G1 and S/G2/M, respectively, but with an EMGD, which must consist of the exponent included in the G1 distribution, and of the gamma distribution summing-up the processes captured by the non-exponential component of G1 distribution and by the S/G2/M distribution. The resulting estimates of τ for G1 and IDT, which are shown in Fig. 5 under distribution curves, suggest that there is a correspondence between G1 and IDT in this regard, as should be expected, although the correspondence may be compromised by the noisiness of input data.

Taken together, the data shown in Fig. 5 are interpretable in terms of TPM (and EMGD by inference) upon not more caveats than it is required to support the stretched model using the same data. As a descriptive tool, EMGD is better than the lognormal or Wald distribution used for all phases in the stretched model and than Wald suggested for G1 phase in the continnum model.

Two essential advantages of the stretched model over TPM are claimed in (Dowling et al., 2014). One is that the former is more consistent with the fact that the variabilities of G1 and the other phases of the cell cycle are comparable. The other is that the durations of different phases of the cell cycle and of defined phases of the cell cycle of different cells in a lineage are correlated.

However, using EMGD (as well as EMG) to treat IDT data in terms of TPM shows that the principal difference between the probabilistic and the deterministic part of the cell cycle is not in that the former is highly variable whereas the latter is not, but in the modes of their variabilities: the former conforms to the exponential, whereas the latter, to the gamma distribution, which may be blurred toward the Gaussian. The variance of the gamma distribution may well be as high as or even higher than that of the exponential one.

The part A may, however, be regarded as more variable in the sense that its mean duration is more sensitive to different intracellular and environmental cues (see 3.2) strengthening its importance as a determinant of the proportion of alternative cell fates in a cell population (Golubev, 2010a, 2010b, 2012a, 2012b) and, correspondingly, the importance of the assessment of changes in its duration, which is possible with EMGD or EMG but not with other suggested models of IDT distributions.

As to correlations, although the original TPM does not predict them, it does not forbid them. Plausible factors that can exert unidirectional influences on all parts or phases of the cell cycle may include, but not limited to: temperature; the availability of ATP or reducing equivalents needed for processes involved in rate limiting of whatever occurs in the cell; the strength of a signal influencing all parts of the cell cycle, e.g., via c-myc expression; and the involvement of si-RNA in the cell cycle. This all may make grounds for TPM elaboration, but not necessarily for its rejection.

4. EMGD compared with other functions used to fit time distributions related to gene transcription and protein expression

The activity of some genes studied in sufficient details has been shown to consist of irregularly alternating periods of the presence or absence of PolII molecules involved in mRNA production at the transcribed portions of the genes (Harper et al., 2011; Suter et al., 2011a, 2011b; Yunger et al., 2010). Such discontinuous mode of mRNA production, which is manifested as transcriptional bursts, makes a basis to regard active genes as going through transcriptional cycles consisting of at least two consecutive periods, which were termed as idle and engaged in (Golubev, 2012a).

The emerging consensus is that the durations of the productive (engaged) period are distributed exponentially, consistent with that each such period is terminated by a single event of dissociation or inactivation of moieties required for recruiting PolII molecules to gene promoters. As to the nonproductive (idle) periods needed for the reinitiation of gene promoter ability to recruit PolII, the issue of their duration is more confusing. When the random telegraph model (RTM) was used to derive engaged and idle period durations from the variation of short-lived reporter protein levels in single cells, each of the resulting idle time distributions of several different promoters was interpreted as having a peak and was approximated with the convolution of two exponents (exponentially modified exponent, EME) (Suter et al., 2011a). In their subsequent study focused on connective tissue growth factor (CTGF) promoter, the same authors used the gamma distribution for this purpose (Molina et al., 2013). When direct visualization of nascent mRNA at defined genes in single cells was used, idle time distributions were found to be exponential (Larson et al., 2013; Ochiai et al., 2014). In the study (Cohen et al., 2009) where times to the onsets of protein expression in single cells were determined, the distributions of times termed here as idle were assumed as exponential, although the histograms showing the experimentally obtained distributions may suggest that they are peaked.

Table 8 compares EMGD, EME, and gamma and exponential distributions as approximations to data on the idle time, which

Table 8
Approximation of transcription cycle idle time distributions with different functions.
Data source: online supplementary materials to Suter et al. (2011a).

Gene promoter*	Fit ranks by r^2						Fit ranks by f-statistic						Numeric parameter estimates							
	EMGD	Gamma	Lognorm	EME	Wald	Exp	Gamma	Lognorm	EME	Exp	Wald	EMGD	EMGD			EME		Gamma		Exp
													c	b	τ	τ_1	τ_2	c	b	τ
CTGF	1	2	5	4	6	3	2	5	3	**1**	6	4	**0.13**	2.79	2.81	***3.55***	0.01	1.13	2.81	***3.55***
Sh3kbp1	2	1	5	3	6	4	2	5	3	**1**	6	4	**0.27**	1.22	1.23	***1.70***	0.07	1.27	1.22	***1.90***
DBP	1	2	4	3	5	6	2	4	3	**1**	6	5	**0.29**	1.24	5.35	***6.13***	0.75	1.23	4.37	***6.83***
Serpin1	1	2	3	5	6	4	2	3	4	**1**	6	5	**0.83**	0.22	2.83	***3.12***	0.01	1.21	2.33	***3.12***
Prl2C2	1	3	4	2	6	5	2	3	1	5	6	4	1.27	0.14	0.74	0.70	0.19	1.81	0.45	1.10
Glutaminase	1	2	3	4	5	6	1	2	3	6	4	5	1.33	0.68	0.69	0.83	0.83	2.23	0.68	2.39
Hmga2	1	2	3	4	5	6	1	2	3	6	4	5	1.54	0.84	3.51	2.56	1.97	2.15	2.04	5.99
Bmal1a	1	2	3	4	5	6	1	2	3	6	4	5	1.76	0.73	0.74	1.10	1.10	2.76	0.73	–
Plectin1	2	3	1	4	5	6	2	1	3	5	6	4	2.26	0.35	2.25	2.34	1.12	0.99	1.98	4.49
Bmal1b	1	3	2	5	4	6	2	1	4	6	3	5	3.26	0.23	1.21	1.01	1.01	3.03	0.59	2.93
Per2het	1	3	6	2	4	5	2	6	1	5	3	4	11.45	0.16	7.56	5.30	2.89	2.31	3.18	16.72
H2	1	5	2	3	4	6	4	1	2	6	3	5	17.42	0.03	2.20	1.83	0.52	1.88	1.14	2.76
Per2hom	1	5	2	4	3	6	4	1	3	5	2	6	29.51	0.01	5.65	4.98	1.11	1.88	2.85	7.78
Mean rank:	**1.15**	**2.69**	**3.31**	**3.62**	**4.92**	**5.31**	**2.08**	**2.77**	**2.77**	**4.15**	**4.54**	**4.69**								

* see Suter et al. (2011a) and the main text for the meaning of abbreviations.

were presented as histograms in online supplementary materials to (Suter et al., 2011a). Two other functions added to Table 8 are the lognormal and Wald distributions. EMG and shifted gamma and Weibull distributions, which were included in the panel applied to IDT, are not used in Table 8 because their best-fit forms regularly extend to the negative time domain (negative τ in EMG or negatively shifted gamma and Weibull distributions).

The same reasoning as used in Section 3.1 for to IDT suggests that EMGD is the best descriptive function for the idle time distributions of transcription cycle. The sign test employed in the same way as in 3.1, even though with a smaller sample of datasets, proves that, according to r^2, the EMGD outperforms its nearest rival, the gamma distribution, at $p < 0.01$ (the critical numbers of minuses at $n=13$ are unity for $p < 0.01$ and two for $p < 0.05$).

Moreover, using EMGD may suggest some mechanistic insights. First, the estimates of c for gene idle time distributions are markedly lower than for IDT distributions. The obvious interpretation that fewer intracellular events are required to prepare a gene for transcription than to prepare a cell for passing its RP lends support to the appropriateness of EMGD as a model of distributions of times between different stages of processes observed in cell populations. Parenthetically, at c as low as presented in Table 8, EMG is unsuitable as an approximation of EMGD, at difference from the suitability of EMG as a substitute for EMGD in the analysis of IDT.

Another insight follows from the observation that the estimates of EMGD parameters distinguish a group of four datasets (Table 8), whose c values are below unity and thus are physically senseless. Conspicuously, the same datasets are peculiar in that, according to the f-statistic, they are best fit with exponential distributions, in sharp contrast to other datasets for which the exponent is the worst approximation by either criterion. Another peculiar feature of these four datasets is that treating each of them with EME yields two very different parameters, one closely matching the estimated characteristic time of the respective pure exponent, whereas the other being much smaller. Any of the other datasets has far less discordant EME parameters.

Taken together, the above observations suggest that there exist two classes of idle time distributions, the first one being exponential (each idle period includes a single event, which terminates the period by enabling gene promoter to recruit PolII), and the second one being peaked (each idle period includes additional event(s) needed to prepare gene promoter for being able to recruit PolII). The difference between fit patterns exhibited by the two types is illustrated in Fig. 6.

While the present paper was under revision, a similar conclusion was published by Zoller et al. (2015) based on a mechanistic model applied to experimental results reported in Suter et al. (2011a). Zoller et al. (2015) used Prlc2 gene and glutaminase gene to exemplify, respectively, the exponential (class 1) and peaked (class 2) idle time distributions. According to Table 8 of the present paper, glutaminase gene is attributed to as class 1, too. As to Prlc2, although it is attributed to class 2, it has the smallest c among all class 2 cases. It thus may be treated as an apparently borderline case, which may be attributed to either class depending on subtle differences in the estimates of c obtained with different procedures applied to noisy data.

Besides promoters presented in the upper four lines in Table 8, promoters studied by direct visualization of nascent mRNA at respective genes, i.e. NANOG (Ochiai et al., 2014) and ecdysone-responsive promoters (Larson et al., 2013), may be attributable to class 2 as follows from treating histograms reported in these two papers in the same way as in the case of datasets presented in Table 8 (not shown).

Data on the distributions of times to protein appearance in single cells (Cohen et al., 2009) actually relate to times of reaching of arbitrary low defined protein levels. Such distributions are generated by processes, which include, besides the initiation of nascent mRNA formation, a series of other processes, including mRNA processing, transport and translation and protein accumulation to a defined level. The overall time of passing all these steps may be gamma-distributed making the overall distribution of times to protein appearance be distributed according to EMGD irrespective of whether idle time distribution are exponential or peaked. Indeed, EMGD was found to be the best-fit function (Fig. 7) for histograms presented in Cohen et al. (2009).

5. Discussion

The analysis presented in this paper suggests that EMGD is the best descriptive function for IDT distributions and for a subset of idle time distributions of the transcription cycle, and that the applicability of EMGD to IDT distributions is more than formal being consistent with the transition probability model of the cell cycle.

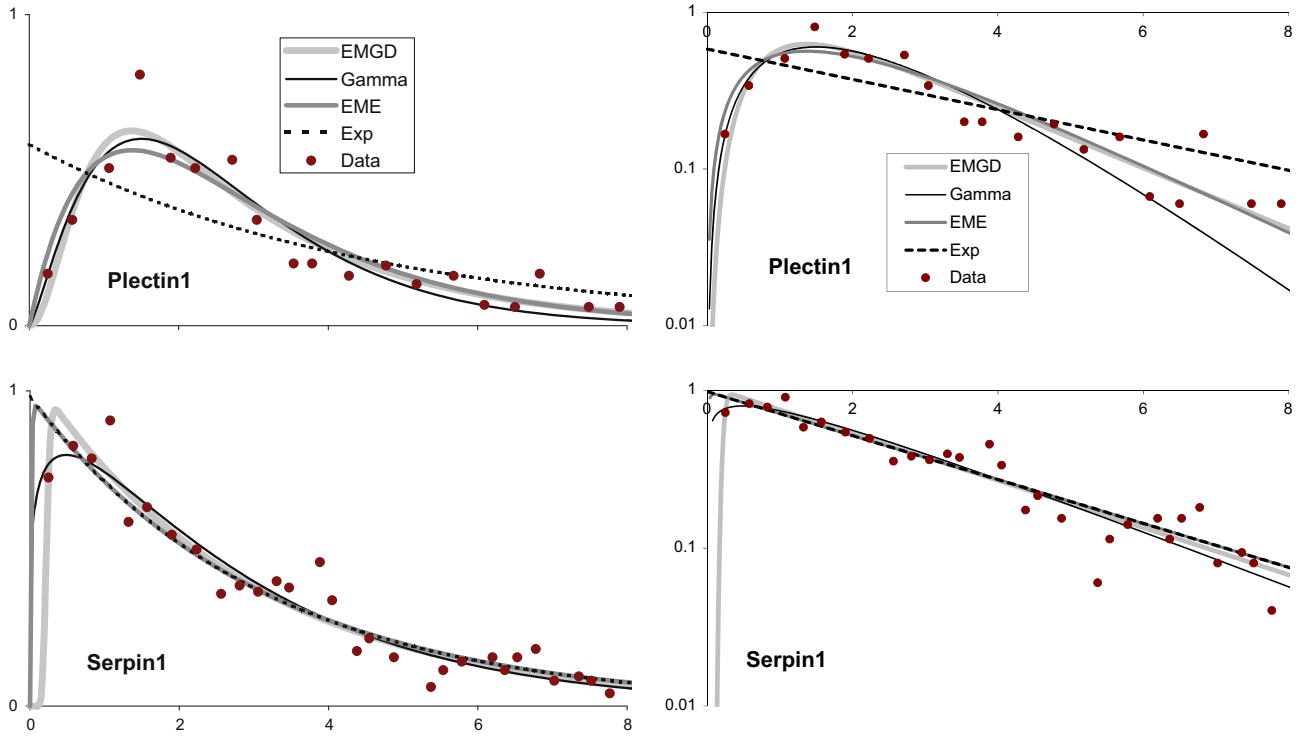

Fig. 6. Two types of distributions of the idle period of transcription cycle. Upper plots: the distribution is most likely peaked. Lower plots: the distribution is most likely exponential. On the right are semilogarithmic transforms of the on the left. Data are derived from Suter et al. (2011a).

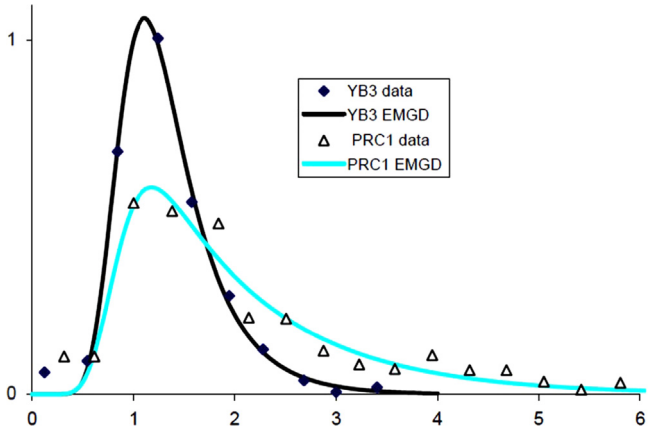

Fig. 7. Distributions of times (hours) to the onsets of YB3 and PRC1 proteins detectability in single cells after cell entry into the cell cycle. Data are derived from Cohen et al. (2009).

Now what if, while having grounds to believe that EMGD or other function is preferable based on theoretical premises and/or extended experience of applying it to a certain class of data, one sees that this function, compared with alternatives, fails to provide the best fits to particular cases, such as exemplified with some entries in Tables 4 through 8? Then the cases are more likely to represent the effects or some uncontrolled factors and, if found on a regular bases, to deserve further investigation or appropriate elaboration of the experimental procedure or model used, rather than to suggest that they are attributable to another class of phenomena or that the model should be rejected altogether. In particular, multiple subtle confounding factors may cause a sort of Gaussian blur of any peak function convoluted with an exponent. The result will be that, in practice, EMG is not worse than EMGD in approximating the shapes of IDT distributions whose means are much greater than their standard deviations. However, EMG is inappropriate when the two parameters are comparable, such as in distributions related to gene transcription times.

Another biological discipline where EMGD may find applicability is psychophysics where exponentially modified Gaussian, termed ex-Gaussian in this field, is used to approximate response time distributions (Palmer et al., 2011). Checking some of published response time distributions using the same procedure as applied here to IDT shows that, in the lists of functions ranked according to their fit to experimental data, EMGD stands next to EMG, which usually occupies the first position (unpublished observations). In this case, too, EMG superiority may be caused by the Gaussian blur of the gamma distribution component convolved with the exponential component of the overall distribution.

The appropriateness of exponentially modified peak functions to the variabilities of time related to decision making at the cellular and organismal levels may reflect a general biological decision-making strategy. Decisions are chosen by random events, which occur in relatively long stochastically varying time intervals and thus constitute a low-rate process in a population of decision-making entities. The probability of a particular choice depends on conditions. The chosen decision is manifested after having been executed via a series of predetermined events occurring in relatively short intervals and thus constituting processes having high rate constants and captured by functions yielding a peaked distribution upon their convolution.

Time distributions reflecting such strategy and approximated with EMGD include IDT distributions. Assuming that their shapes are determined by the arrangement of processes involved in the cell cycle, the conformance of such shapes to EMGD is consistent with that at least one of processes implicated in cells passage through the cell cycle has a very low rate constant compared with all other processes.

In fact, the main issue addressed here is not whether gamma, normal or other distribution is the best approximation of a peak convolved with the exponential component of an entire IDT distribution, but whether an exponent whose characteristic time is

comparable with the mean IDT is involved in shaping the entire IDT distribution. If it is involved, then IDT distributions are generally consistent with the transition probability model (TPM). The common feature of all alternatives suggested for IDT distributions is to reject this involvement and, by inference, the TPM.

An IDT distribution is a feature of a cell population. At the level of individual cells in the population, the conformance of its IDT distribution to EMGD means that its cells dwell for a relatively long mean time in the state associated with the process having the lowest rate constant. By "relatively long" it is implied that the stochastically varying times of dwelling in this state may be comparable with the mean IDT. It has been argued (Golubev, 1996; 2010a, 2010b; 2012a, 2012b) that this state corresponds to the restriction point (RP) of the cell cycle, that is, the moments when cells exit their RPs fall within the probabilistic part A of the cell cycle, which is mapped to the G1 phase. RP is suggested to be the state where cells make decisions whether to go on with proliferation or to turn to other options, including differentiation, senescence and apoptosis. Each decision is associated with a specific stochastic event, which by chance happens to be the first to occur. Each such event is suggested to be associated with the formation of a combination of transcription factors at the promoter of gene involved in the actuation of the execution of the chosen option. Intracellular and environmental cues can influence the frequencies of such competing events, and an event occurring at a higher frequency in a cell population (having smaller mean inter-event periods) is more likely to be (but not necessarily is) the first to happen in a specific cell.

The decision of a cell to divide is executed after the cell has exited its RP and goes through the subsequent steps of the cell cycle, which are included in the deterministic part B. This part also comprises processes that precede the RP and prepare cells to enter it. The difference between the two parts, A and B, is not in that the former is highly variable, whereas the latter is not so (its apparent variability being a mere scatter usual for any quantitative biological trait), but in the modes of variability: the former is defined in a cell population by relatively infrequent random intracellular events, which form the exponential component of IDT distributions, whereas the latter, by a series of processes, each constituted by frequently occurring events, which altogether form the peak component of IDT distribution. At the first approximation, the resulting variability of part B may be assumed to correspond to a Gaussian distribution. In the present paper, part B variability is suggested to be more consistent with a gamma distribution, which may be convolved with an exponent whose τ is greater than the scale parameter b of the gamma distribution. However, multiple confounding factors are able to cause a sort of Gaussian blur of any peak function convolved with an exponent. The result is that, in practice, EMG is generally no worse than EMGD in approximating the shapes of IDT.

In the present paper, the two exponentially modified peak functions, EMG and EMGD, are shown to outperform in a systematic and statistically significant manner other positively skewed functions in the ability to approximate IDT distributions. Both of them do not show such superiority when applied to other positively skewed distributions, such as of cell size, protein levels in single cells, or telomere length (not shown). Therefore, the adequacy of EMG and EMGD to IDT is specific rather than attributable to their enhanced ability to fit everything due to additional adjustable parameters. Because for both, EMG and EMGD, closed-form formulae are available, they may be used to derive the mean time of cell dwelling in the probabilistic part of the cell cycle from IDT data. These estimates are in the range of 0.5–5 h. Under influences, such as anticancer drugs, that can interfere with the cell cycle, the time may increase up to 10 h, making a longer window to occur for events associated with alternative options, such as differentiation or apoptosis (Golubev, 2010b, 2012b).

A fundamental question arises: what is the nature of the random event (or a specific constellation of events, which is an event in its own right) that determines cell exit from RP and, in a single cell, can happen at any time, up to several hours, after the cell has reached its RP?

The mainstream approach in tackling this issue is based on the assumption that protein concentrations, enzyme activities, and gene expression levels change continuously, depending on interactions between genes, enzymes and other proteins. The interactions are described with ordinary differential equations (ODE), which operate with the rate constants of interactions between the populations of entities included in a model (e.g. Conradie et al., 2010; Pfeuty, 2012). When a postulated pattern of interactions includes feedforward and/or positive feedback loops, the whole system of interacting populations may exhibit dynamic switches between two or more relatively stable states suggested by the solutions of ODE systems describing the interactions. One such study (Yao et al., 2008) designed to check the validity of a system, which includes Rb, c-myc, E2F, cyclin E and cdk2, suggested that cell passage through RP occurs by switches between two states characterized by low (before RP) and high (after RP) activities of genes responsive to E2F transcription factors, which, when Rb is hyperphosphorylated, are released from inhibition produced by hypophosphorylated Rb. To make the switches irregular, a stochastic component accounting of low, and thus fluctuating, concentrations of system's components was introduced by making use of the Gillespie algorithm (Lee et al., 2010).

The fundamental flaw of such approach is that, according to it, processes observed in cell populations are treated as if determined by interacting populations of molecular species distributed in a common ideally mixed medium and characterized by continuous changes, including fluctuations, in their concentrations. However, the key determinants of the course of cell cycle, i.e. proteins, are confined to separate cells at quantities, which significantly differ between their receptacles. The ODE approach, which is picked up from modeling of metabolic pathways involving genuine molecular populations highly populated even at the single-cell level, is hardly suitable to gene expression, at minimum because genes are present in a cell in one or a few copies embedded in highly crowded environments. This is enough to consider the notion of concentration as senseless with regard to genes.

No less importantly, the ability of a system described with ordinary or stochastic differential equations to undergo switches is the property of the system as a whole, but not of any of its component taken separately. This means that the behavior of any such model is crucially dependent on its structure and on the assumptions about the concentrations of its components and about the rate constants of their interactions. Meanwhile, the ideas about the possible structures of such systems are constantly changing as more components are experimentally found to be important. For example, si-RNA could not be included in cell cycle models before si-RNA were discovered and their involvement in the cell cycle was recognized (Yan et al., 2012). Therefore, any such model is based on assumptions that cannot but be highly arbitrary. It is important in this regard that, even when the pivotal components of ODE-based models, such as cyclins or cyclin-dependent kinases, are knocked-out, the resulting knockout cells are still capable of going through the cell cycle (Li et al., 2009). For all those reasons, discrete-event models admitting the possibility of spontaneous changes in some of their components seem to be more adequate. Indeed, it has been shown (Golubev, 2010b) that even a toy model of this sort can reproduce the experimental findings reported in (Yao et al., 2008) as well as the elaborate ODE model used in this paper does.

The nature of such discrete events may relate to the recognized fact that the activities of a number of genes consist of spontaneously alternating irregular periods of the presence and absence of RNA polymerase molecules at the transcribed part of a gene (see references in Section 4 of this paper). The durations of the periods of gene engagement in transcription are distributed exponentially, consistent with the assumption that such periods are terminated by the dissociation of a pre-initiating complex of transcription factors from gene promoter. A similar kinetics may be envisioned for the idle periods, i.e. times to the onsets of gene expression, in the case of dissociation of a suppressor of transcription from gene promoter. Treating reported idle time distributions with EMGD (Section 4) suggests that, indeed, some of them are likely to be exponential. Their characteristic times may range from 1 to 6 h, that is, may be comparable with the mean duration of the probabilistic part A of the cell cycle, consistent with the idea that passing the restriction point of the cell cycle may be driven by sporadic transcriptional events committing cells to different competing options, such as further proliferation, or differentiation, or apoptosis.

In an attempt to consider these events in possibly more detail (Golubev, 2010b), they were associated with what might occur at CycE promoter upon pocket-protein phosphorylation. However, this is the maximal level of certainty possible with evidence available so far about the molecular mechanisms that may generate the exponential component of IDT variabilities. Whatever the critical event, the distribution of times between its occurrence and cell passage through the restriction point of the cell cycle is captured by the exponential decay function convolved with a peak function to yield the entire IDT distribution, which may be approximated with EMGD.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jtbi.2015.12.027.

References

Avigad, G., 2008. Computers in mathematical enquiry. In: Mancusco, P. (Ed.), The Phylosophy of Mathematical Practice. Oxford Univercity Press, Oxford, pp. 302–316.

Castor, L.N., 1980. A G1 rate model accounts for cell-cycle kinetics attributed to "transition probability". Nature 287, 857–859.

Cohen, A.A., Kalisky, T., Mayo, A., Geva-Zatorsky, N., Danon, T., Issaeva, I., Kopito, R.B., Perzov, N., Milo, R., Sigal, A., Alon, U., 2009. Protein dynamics in individual human cells: Experiment and theory. PloS One 4, e4901. http://dx.doi.org/10.1371/journal.pone.0004901.

Conradie, R., Bruggeman, F.J., Ciliberto, A., Csikasz-Nagy, A., Novak, B., Westerhoff, H.V., Snoep, J.L., 2010. Restriction point control of the mammalian cell cycle via the cyclin E/Cdk2:p27 complex. FEBS J. 277, 357–367.

Cooper, S., 1982. The continuum model: statistical implications. J. Theor. Biol. 94, 783–800.

Cooper, S., 2015. Gene expression during the cell cycle: obfuscation of original cell-cycle gene expression data by normalization. J. Cells 1, 1–7.

Dixon, W.J., Mood, A.M., 1946. The statistical sign test. J. Am. Stat. Assoc. 41, 557–566.

Dowling, M.R., Kan, A., Heinzel, S., Zhou, J.H.S., Marchingo, J.M., Wellard, C.J., Markham, J.F., Hodgkin, P.D., 2014. Stretched cell cycle model for proliferating lymphocytes. Proc. Natl. Acad. Sci. USA 111, 6377–6382. http://dx.doi.org/10.1073/pnas.1322420111.

Duffy, K.R., Wellard, C.J., Markham, J.F., Zhou, J.H.S., Holmberg, R., Hawkins, E.D., Hasbold, J., Dowling, M.R., Hodgkin, P.D., 2012. Activation-induced B cell fates are selected by intracellular stochastic competition. Science 335, 338–341. http://dx.doi.org/10.1126/science.1213230.

Floyd, D.L., Harrison, S.C., van Oijen, A.M., 2010. Analysis of kinetic intermediates in single-particle dwell-time distributions. Biophys. J. 99, 360–366. http://dx.doi.org/10.1016/j.bpj.2010.04.049.

Gabriel, P., Garbett, S.P., Quaranta, V., Tyson, D.R., Webb, G.F., 2012. The contribution of age structure to cell population responses to targeted therapeutics. J. Theor. Biol. 311, 19–27. http://dx.doi.org/10.1016/j.jtbi.2012.07.001.

Geiler-Samerotte, K.A., Bauer, C.R., Li, S., Ziv, N., Gresham, D., Siegal, M.L., 2013. The details in the distributions: why and how to study phenotypic variability. Curr. Opin. Biotechnol. 24, 752–759. http://dx.doi.org/10.1016/j.copbio.2013.03.010.

Golubev, A., 2010a. Exponentially modified Gaussian (EMG) relevance to distributions related to cell proliferation and differentiation. J. Theor. Biol. 262, 257–266.

Golubev, A., 2010b. Random discrete competing events vs. dynamic bistable switches in cell proliferation in differentiation. J. Theor. Biol. 267, 341–354.

Golubev, A., 2012a. Genes at work in random bouts. BioEssays 34, 311–319.

Golubev, A., 2012b. Transition probability in cell proliferation, stochasticity in cell differentiation, and the restriction point of the cell cycle in one package. Progr. Biophys. Mol. Biol. 110, 87–96.

Golubev, A.G., 1996. Accidental necessity, initiation of transcription, cell differentiation, and necessary accidents. Biochemistry 61, 928–938.

Gomes, F.L.A.F., Zhang, G., Carbonell, F., Correa, J.A., Harris, W.A., Simons, B.D., Cayouette, M., 2011. Reconstruction of rat retinal progenitor cell lineages in vitro reveals a surprising degree of stochasticity in cell fate decisions. Development 138, 227–235. http://dx.doi.org/10.1242/dev.059683.

Harper, C.V., Finkenstedt, B., Woodcock, D.J., Friedrichsen, S., Semprini, S., Ashall, L., Spiller, D.G., Mullins, J.J., Rand, D.A., Davis, J.R.E., White, M.R.H., 2011. Dynamic analysis of stochastic transcription cycles. PLOS Biol. 9, e1000607.

Hawkins, E.D., Markham, J.F., McGuinness, L.P., Hodgkin, P.D., 2009. A single-cell pedigree analysis of alternative stochastic lymphocyte fates. Proc. Natl. Acad. Sci. USA 106, 13457–13462. http://dx.doi.org/10.1073/pnas.0905629106.

Hawkins, E.D., Turner, M.L., Dowling, M.R., van Gend, C., Hodgkin, P.D., 2007. A model of immune regulation as a consequence of randomized lymphocyte division and death times. Proc. Natl. Acad. Sci. USA 104, 5032–5037. http://dx.doi.org/10.1073/pnas.0700026104.

Holzel, M., Kohlhuber, F., Schlosser, I., Holzel, D., Luscher, B., Eick, D., 2001. Myc/Max/Mad regulate the frequency but not the duration of productive cell cycles. EMBO Rep. 2, 1125–1132. http://dx.doi.org/10.1093/embo-reports/kve251.

Hurwitz, C., Tolmach, L.J., 1969. Time-lapse cinemicrographic studies of X-irradiated HeLa S3 cells. Biophys. J. 9, 607–633. http://dx.doi.org/10.1016/S0006-3495(69)86407-6.

Lainey, E., Wolfromm, A., Sukkurwala, A.Q., Micol, J.-B., Fenaux, P., Galluzzi, L., Kepp, O., Kroemer, G., 2013. EGFR inhibitors exacerbate differentiation and cell cycle arrest induced by retinoic acid and vitamin D3 in acute myeloid leukemia cells. Cell Cycle 12, 2978–2991. http://dx.doi.org/10.4161/cc.26016.

Larson, D.R., Fritzsch, C., Sun, L., Meng, X., Lawrence, D.S., Singer, R.H., 2013. Direct observation of frequency modulated transcription in single cells using light activation. eLife 2. http://dx.doi.org/10.7554/eLife.00750.

Leander, R., Allen, E.J., Garbett, S.P., Tyson, D.R., Quaranta, V., 2014. Derivation and experimental comparison of cell-division probability densities. J. Theor. Biol. 359, 129–135. http://dx.doi.org/10.1016/j.jtbi.2014.06.004.

Lee, H., Perelson, A., 2008. Modeling T Cell proliferation and death in vitro based on labeling data: generalizations of the Smith–Martin cell cycle model. Bull. Math. Biol. 70, 21–44.

Lee, T.J., Yao, G., Bennett, D.C., Nevins, J.R., You, L., 2010. Stochastic E2F activation and reconciliation of phenomenological cell-cycle models. PLOS Biol. 8, e1000488.

León, K., Faro, J., Carneiro, J., 2004. A general mathematical framework to model generation structure in a population of asynchronously dividing cells. J. Theor. Biol. 229, 455–476. http://dx.doi.org/10.1016/j.jtbi.2004.04.011.

Li, W., Kotoshiba, S., Kaldis, P., 2009. Genetic mouse models to investigate cell cycle regulation. Transgen. Res. 18, 491–498. http://dx.doi.org/10.1007/s11248-009-9276-x.

Maler, A., Lutscher, F., 2010. Cell-cycle times and the tumour control probability. Mathemat. Med. Biol. 27, 313–342. http://dx.doi.org/10.1093/imammb/dqp024.

Molina, N., Suter, D.M., Cannavo, R., Zoller, B., Gotic, I., Naef, F., 2013. Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. Proc. Natl. Acad. Sci. USA 110, 20563–20568. http://dx.doi.org/10.1073/pnas.1312310110.

Nakaoka, S., Inaba, H., 2014. Demographic modeling of transient amplifying cell population growth. Math. Biosci. Eng. 11, 363–384. http://dx.doi.org/10.3934/mbe.2014.11.363.

Ochiai, H., Sugawara, T., Sakuma, T., Yamamoto, T., 2014. Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. Sci. Rep. 4. http://dx.doi.org/10.1038/srep07125.

Palmer, E.M., Horowitz, T.S., Torralba, A., Wolfe, J.M., 2011. What are the shapes of response time distributions in visual search? J. Exp. Psychol.: Hum. Percept. Perform. 37, 58–71. http://dx.doi.org/10.1037/a0020747.

Pfeuty, B., 2012. Strategic cell-cycle regulatory features that provide mammalian cells with tunable G1 length and reversible G1 arrest. PloS One 7, e35291.

Schwabe, A., Rybakova, KatjaB N., Bruggeman, FrankB J., 2012. Transcription stochasticity of complex gene regulation models. Biophys. J. 103, 1152–1161. http://dx.doi.org/10.1016/j.bpj.2012.07.011.

Slocum, H.K., Parsons, J.C., Winslow, E.O., Broderick, L., Minderman, H., Tóth, K., Greco, W.R., Rustum, Y.M., 2000. Time-lapse video reveals immediate heterogeneity and heritable damage among human ileocecal carcinoma HCT-8 cells treated with raltitrexed (ZD1694). Cytometry 41, 252–260. http://dx.doi.org/10.1002/1097-0320(20001201)41:4<252::AID-CYTO3>3.0.CO;2-X.

Smith, J.A., Martin, L., 1973. Do cells cycle? Proc. Natl. Acad. Sci. USA 70, 1263–1267.

Stukalin, E.B., Aifuwa, I., Kim, J.S., Wirtz, D., Sun, S.X., 2013. Age-dependent stochastic models for understanding population fluctuations in continuously cultured cells. J. R. Soc. Interface 10. http://dx.doi.org/10.1098/rsif.2013.0325.

Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., Naef, F., 2011a. Mammalian genes are transcribed with widely different bursting kinetics. Science 332, 472–474.

Suter, D.M., Molina, N., Naef, F., Schibler, U., 2011b. Origins and consequences of transcriptional discontinuity. Curr. Opin. Cell Biol. 23, 657–662.

Tyson, D.R., Garbett, S.P., Frick, P.L., Quaranta, V., 2012. Fractional proliferation: a method to deconvolve cell population dynamics from single-cell data. Nat. Methods 9, 923–928. http://dx.doi.org/10.1038/nmeth.2138.

Weber, T.S., Jaehnert, I., Schichor, C., Or-Guil, M., Carneiro, J., 2014. Quantifying the length and variance of the eukaryotic cell cycle phases by a stochastic model and dual nucleoside pulse labelling. PLOS Comput. Biol. 10, e1003616. http://dx.doi.org/10.1371/journal.pcbi.1003616.

Yan, F., Liu, H., Hao, J., Liu, Z., 2012. Dynamical behaviors of Rb-E2F Pathway including negative feedback loops involving miR449. PloS One 7, e43908. http://dx.doi.org/10.1371/journal.pone.0043908.

Yao, G., Lee, T.J., Mori, S., Nevins, J.R., You, L., 2008. A bistable Rb-E2F switch underlies the restriction point. Nat. Cell Biol. 10, 476–482.

Yunger, S., Rosenfeld, L., Garini, Y., Shav-Tal, Y., 2010. Single-allele analysis of transcription kinetics in living mammalian cells. Nat. Methods 7, 631–633.

Zilman, A., Ganusov, V.V., Perelson, A.S., 2010. Stochastic models of lymphocyte proliferation and death. PloS One 5, e12775.

Zoller, B., Nicolas, D., Molina, N., Naef, F., 2015. Structure of silent transcription intervals and noise characteristics of mammalian genes. Mol. Syst. Biol. 11, 823. http://dx.doi.org/10.15252/msb.20156257.